AD-A253 622

ITATION PAGE

Form Approved
OMB No. 0704-0188

ed to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, viewing the collection of information. Send comments regarding this burden estimate or any other aspect of this irden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson ffice of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
June 1992

3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
The Total Time on Test Statistic and Age-Dependent Censoring

5. FUNDING NUMBERS
DAAL03-91-G-0046

6. AUTHOR(S)
Dr. Roger M. Cooke

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The Regents of the University of California
c/o Sponsored Projects Office
336 Sproul
University of California
Berkeley, CA 94720

8. PERFORMING ORGANIZATION REPORT NUMBER

DTIC
S AUG 06 1992 D

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARO 27993.2-MA

11. SUPPLEMENTARY NOTES
The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

It is known that observations from any censored life (or competing risk) process can be described by a random censoring (or independent risk) model. On the other hand it is impossible to verify that the censoring is really random. A class of age-dependent censoring processes are defined and the class of corresponding subsurvival functions is described. Exponential life variables censored by independent variables can also be described by age-dependent censoring models if the censoring variable is DFR. The total time on test statistic consistently estimates the expected life of an exponential life variable under random right censoring. If the censoring is age-dependent, the Total Time on Test statistic will severely overestimate the expected life of the variable of interest. Care should be taken to motivate the application of the total time on test statistic in such situations.

92 7 1 141

92-20856

14. SUBJECT TERMS
Total time on test statistic; subsurvival function; censoring; non-random censoring; competing risk; failure rate; IFR, DFR.

15. NUMBER OF PAGES
15

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

THE TOTAL TIME ON TEST STATISTIC AND AGE-DEPENDENT CENSORING

Interim Technical Report

by

Dr. Roger M. Cooke

June 1992

U.S. Army Research Office

Grant Number: DAAL03-91-G-0046

University of California at Berkeley

The Total Time on Test Statistic and Age-Dependent Censoring

Roger M. Cooke *

June 17, 1992

Abstract

It is known that observations from any censored life (or competing risk) process can be described by a random censoring (or independent risk) model. On the other hand it is impossible to verify that the censoring is really random. A class of age-dependent censoring processes are defined and the class of corresponding subsurvival functions is described. Exponential life variables censored by independent variables can also be described by age-dependent censoring models if the censoring variable is *DFR*. The total time on test statistic consistently estimates the expected life of an exponential life variable under random right censoring. If the censoring is age-dependent, the Total Time on Test statistic will severly overestimate the expected life of the variable of interest. Care should be taken to motivate the application of the total time on test statistic in such situations.

Key Words: Total time on test statistic, subsurvival function, censoring, non-random censoring, competing risk, failure rate, IFR, DFR.

*Department of Mathematics and Informatics, Delft University of Technology, Delft, The Netherlands. This research was partially supported by a grant to the University of California at Berkeley by the Army Research Office. This paper was written while the author was on sabbatical leave at the Department of Industrial Engineering and Operations Research, University of California at Berkeley

1 Introduction

For exponential reliability life variables subjected to random right censoring, the *Total Time on Test* statistic

$$TTT = \frac{\text{total time on test}}{\#\text{ of failures}} \tag{1}$$

is the inverse maximum likelihood estimator of the failure rate. It is widely used in reliability analysis to estimate the expectation and failure rate of the uncensored life process.

In practice it is impossible to determine whether the censoring is really random or whether the uncensored life process is really exponential[1]. In many situations, other assumptions would also be prime facie plausible. In practice, analysts regard the assumptions underlying the TTT as idealizations, and expect that the error caused by using these idealizations will be small.

This paper shows that the idealization is not harmless. Two classes of censoring models for right censored life variables are considered. The first class is random right censoring. The second class assumes that *whether* a life variable is censored is independent of its age, but *given* that it is censored, the censoring time may depend on age. These are termed *age-dependent censoring models.* This type of coupling between age and censoring time has prime facie plausibility in many situations. Suppose a reliability component shows symptoms prior to failure. *If* these symptoms are observed by personnel, the component will be taken off line and repaired, hence censored. However, the process of observation may be described by a random variable which is independent of the component's age. In medical cohort studies, patients who feel healthy after a treatment might be more likely to move to another city, thus causing a censored observation, whereas others tend to stay put.

It has been known at least since Cox [4] that any competing risk or censored life process can be described by an independent model (see also [5], [6], [10].) These authors draw attention to the problem of "identifiability", namely that the empirical failure data do not determine a unique model. Empirical failure data determine at most a pair of subsurvival functions.

[1] Tests for exponentiality assuming random censoring can be given [4]

This paper describes the set of subsurvival functions which can be described by age-dependent censoring. A necessary and sufficient condition is that the conditional subsurvival function of the life variable should dominate the conditional subsurvival function of the censoring variable. If the life variable is exponential and if the censoring is random, then the conditional subsurvival function of the life variable dominates that of the censoring variable if the latter has a decreasing failure rate (DFR). The dominance relation is reversed if the censoring variable has an increasing failure rate (IFR). Cases in which the censoring variables are also exponentially distributed lie on the boundary between these two subsets. A concluding section points out that the consequences of choosing the wrong model can be severe, when estimating the expectation ofs the life variable. This is especially true when censoring is frequent, as is usually the case in reliability applications. A "dual" model in which the roles of the life and censoring variables is reversed, is discussed briefly.

2 Notation and Definitions

Let X and Z be positive random varibles with distribution functions F_X and F_Z. $X \perp Z$ says that X and Z are independent, $X \sim Z$ says that $F_X = F_Z$ and $X \wedge Z = \min\{X, Z\}$. $S_X = 1 - F_X$ is the *survival function* of X. $S_X^*(t)$ $=P\{X \geq t \text{ and } X < Z\}$ is the *subsurvival function*. Note that S_X^* depends on Z, although this fact is suppressed in the notation. Note also that $S_X^*(0)$ $=P\{X < Z\}$. We assume that $Y = \{\min\{X, Z\}, 1_{\{X<Z\}}\}$ is observable, where $1_{\{\bullet\}}$ is the indicator function of the event $\{\bullet\}$. X is considered the variable of interest, and Z is the censoring variable. By observing Y we observe the smallest of these two, and observe which it is. It is convenient to speak of variables X_i as *components* and of variables Z_i as *censoring times.* If $Z_i < X_i$ we say the component i is censored. The value assigned to Z_i when $Z_i > X_i$ is arbitrary.

We say that S_1^* and S_2^* form a *subsurvival pair* if

1. S_1^* and S_2^* are non-negative non-increasing real functions with $S_1^*(0) \leq 1$ and $S_2^*(0) \leq 1$.

2. $\lim_{t\to\infty} S_1^*(t) = \lim_{t\to\infty} S_2^*(t) = 0$; and

3. $S_1^*(0) = 1 - S_2^*(0)$.

Let $Y_i; i = 1, 2, \cdots$ be *i.i.d.* copies of Y, and $\mathbf{Y} = Y_1, Y_2, \cdots$. The subsurvival pair S_X^* and S_Z^* determine the distribution of Y and hence contains all the information which can be extracted from observing $\mathbf{Y}$.

3 Two Classes of Models

We study two classes of models for $\mathbf{Y}$.

- *Class I: random right censoring:* $X \perp Z$
- *Class II: age-dependent censoring:* $Z = X - \xi$; $X \perp \mathrm{sgn}\xi$

In age-dependent censoring the probability of a component being censored is independent of its age, but *given* that it is censored, the time at which it is censored may depend on its age. Putting $\sigma = (X - \xi)/X$, we can also express an age dependent censoring variable as $Z = X\sigma$.

In this section we determine under what circumstances right censored life data can distinguish between these two model classes. The following theorem has been in the folklore for many years (see , [1], [3],[4], [6], [7], [10]), but a proof sketch is provided for completeness[2].

Theorem 1 *Let X and Z be continuous life variables.*

1. *If $X \perp Z$. and $X' \perp Z'$ with $S_X^* = S_{X'}^*$ and $S_Z^* = S_{Z'}^*$, then $X \sim X'$ and $Z \sim Z'$.*
2. *if S_1^* and S_1^* are a subsurvival pair and are continuous, then there exist independent random variables X and Z such that $S_X^* = S_1^*$ and $S_Z^* = S_2^*$.*

[2]Most authors show that any censored life process can be modeled by a process for which $P\{X > t \cap Z > t\} = P\{X > t\}P\{Z > t\}$, which is implied by but not equivalent with independence [6]. The second statement in **Theorem 1** is proved by Tsiatis [10]; the version here is simplified by the restriction to two "competing risks".

Proof: 1: Using $S_X S_Z = S_X^* + S_Z^*$ and $dS_X^*(t) = S_Z(t)dS_X(t)$;

$$S_X(t) = exp\left\{\int_0^t \frac{dS_X^*(u)}{S_X^*(u) + S_Z^*(u)}\right\}, \tag{2}$$

$$S_Z(t) = exp\left\{\int_0^t \frac{dS_Z^*(u)}{S_X^*(u) + S_Z^*(u)}\right\}. \tag{3}$$

Similar expressions hold for X' and Z'. For 2, choose $X \perp Z$ with survival functions

$$S_X(t) = exp\left\{\int_0^t \frac{dS_1^*(u)}{S_1^*(u) + S_2^*(u)}\right\}; \tag{4}$$

$$S_Z(t) = exp\left\{\int_0^t \frac{dS_2^*(u)}{S_1^*(u) + S_2^*(u)}\right\}. \tag{5}$$

From 1 we know that these survival functions can also be written as (2) and (3). Thus we have

$$\frac{dS_1^*}{dS_X^*} = \frac{S_1^* + S_2^*}{S_X^* + S_Z^*} = \frac{dS_2^*}{dS_Z^*}.$$

Equating the product of equations (2) and (3) with the product of (4) and (5):

$$\frac{d(S_1^* + S_2^*)}{S_1^* + S_2^*} = \frac{d(S_X^* + S_Z^*)}{S_X^* + S_Z^*} \tag{6}$$

so that $\ln(S_1^* + S_2^*) = \ln(S_X^* + S_Z^*)$, or

$$\frac{S_1^* + S_2^*}{S_X^* + S_Z^*} = 1 = \frac{dS_Z^*}{dS_2} = \frac{dS_X^*}{dS_1}. \tag{7}$$

Since $\lim_{t\to\infty}(S_1^*(t) - S_X^*(t)) = 0$, and it follows that $S_1^* = S_X^*$ and similarly for S_Z^*. $\square$

From Theorem 1, any continuous subsurvival pair defines a unique model of *class I*. To get this type of result for *class II* we must restrict to continuous strictly monotonic subsurvival pairs.

Theorem 2 *Let $\{S_1^*, S_2^*\}$ be a pair of continuous strictly monotonic subsurvival functions; then the following are equivalent:*

1. *There exist random variables ξ and X with $sgn(\xi) \perp X$ such that*

$$S_1^*(t) = P\{X > t \cap \xi < 0\} \tag{8}$$

$$S_2^*(t) = P\{X - \xi > t \cap \xi > 0\} \tag{9}$$

2. *For all $t > 0$*

$$\frac{S_1^*(t)}{S_1^*(0)} > \frac{S_2^*(t)}{S_2^*(0)} \tag{10}$$

Proof: (1) implies (2). Since $X \perp sgn(\xi)$:

$$\begin{aligned} S_1^*(t) &= P\{X > t | \xi < 0\} P\{\xi < 0\} = P\{X > t | \xi < 0\} S_1^*(0) = P\{X > t | \xi > 0\} S_1^*(0) \\ S_2^*(t) &= P\{X - \xi > t | \xi > 0\} P\{\xi > 0\} = P\{X - \xi > t | \xi > 0\} S_2^*(0). \end{aligned}$$

Hence for $t > 0$;

$$\frac{S_1^*(t)}{S_1^*(0)} = P\{X > t | \xi > 0\} > P\{X - \xi > t | \xi > 0\} = \frac{S_2^*(t)}{S_2^*(0)}.$$

(2) implies (1): Let X and Z be random variables with survival functions $S_X(t) = S_1^*(t)/S_1^*(0)$; and $S_Z(t) = S_2^*(t)/S_2^*(0)$. Then $S_Z(t) < S_X(t)$ for all $t > 0$, and S_X^{-1} and S_Z^{-1} exist. Choose a random variable $\delta \perp X$; with $P\{\delta = 1\} = 1 - P\{\delta = 0\} = S_2^*(0)$; and put:

$$\xi = \delta(X - S_Z^{-1} S_X(X)) - (1 - \delta) \tag{11}$$

We have $\{\xi > 0\} = \{\delta = 1\}$; $\{\xi = -1\} = \{\delta = 0\}$. Hence $X \perp sgn(\xi)$ and

$$\begin{aligned} P\{X - \xi > t \cap \xi > 0\} &= P\{X - \xi > t | \xi > 0\} S_2^*(0) = \\ P\{S_Z^{-1} S_X(X) > t\} S_2^*(0) &= P\{X > S_X^{-1} S_Z(t)\} S_2^*(0) = \\ (S_X S_X^{-1} S_Z)(t) S_2^*(0) &= S_Z(t) S_2^*(0) = S_2^*(t), \end{aligned}$$

Also $P\{X > t \cap \xi < 0\} = S_X(t) P\{\xi < 0\} = S_X(t) S_1(0) = S_1^*(t)$.□

Observations of a right censored process $\mathbf{Y}$ yield an empirical subsurvival pair which converges to S_X^* and S_Z^*. These in turn express all we can learn from the empirical process. Theorem (1) says that a unique random censoring model will always describe the process. Theorem (2) says that an age-dependent censoring model will describe the process if (10) holds. Uniqueness in this case is impossible as the distribution of ξ given $\xi < 0$ is undetermined.

Of course, if the inequality in (10) were reversed, then a model could be found in which the roles of X and Z were reversed. In other words the sign of ξ would be independent of the censoring variable Z. Such models would be difficult to interprete in a reliability context as it is difficult to assign meaning to Z when $X < Z$. However, in medical cohort studies this type of model might make sense. Typically the end date of such studies is fixed in advance, but patients are admitted into the study at random times, after undergoing some treatment. The variable Z_i denotes the amount of time that patient i is potentially under observation, that is, the difference between the study termination date and i's entrance date. Let X_i denote the time after treatement at which i dies. It is plausible that *whether* i responds to treatment, and hence does not die within the time of the study, is independent of the date at which i entered the study. Hence we could model $X_i = Z_i - \xi$ with $Z_i \perp sgn(\xi)$. Observations fitting this pattern are reported in [9].

The following theorems relate *class I* to the assumptions of exponentiality for X and/or Z.

Theorem 3 *Let $X \perp Z$ with continuous strictly monotonic distribution functions. Then any two of the following statements imply the others*

1. $S_X(t) = \exp^{-\lambda t}$
2. $S_X^*(t) = \frac{\lambda}{\lambda+\gamma} \exp^{-(\lambda+\gamma)t}$
3. $S_Z^*(t) = \frac{\gamma}{\lambda+\gamma} \exp^{-(\lambda+\gamma)t}$
4. $S_Z(t) = \exp^{-\gamma t}$

Proof: For {1,2} $\Rightarrow$ 3 and {2,3} $\Rightarrow$ 1, substitute the information given into equations (2) and (3). {1,4} $\Rightarrow$ 2 is a direct computation. For {1,3} $\Rightarrow$ 4, put $f_Z(u) = -dS_Z(u)/du$, then

$$S_Z^*(t) = \int_t^\infty \exp^{-\lambda u} f_Z(u)du = \frac{\gamma}{\lambda+\gamma} \exp^{-(\lambda+\gamma)t}$$

Taking derivatives of both sides yields $f_Z(t) = \gamma \exp^{-\gamma t}$, which yields 4. $\square$

Corollary If $X \perp Z$ and $S_X(t) = \exp^{-\lambda t}$, $S_Z(t) = \exp^{-\gamma t}$, then

$$\frac{S_Z^*(t)}{S_Z^*(0)} = \frac{S_X^*(t)}{S_X^*(0)}$$

The above property does not uniquely characterize the exponential distribution [5].

Theorem 4 *Let $X \perp Z$ and suppose that there exists a λ such that for all $T > 0$:*

$$P\{X < T \cap X < Z\} = \lambda E(X \wedge Z) 1_{\{X \wedge Z < T\}}; \tag{12}$$

then $S_X(t) = \exp^{-\lambda t}$. Conversely, if $X \perp Z$ and $S_X(t) = \exp^{-\lambda t}$, then (12) holds.

Proof: Using $S_{X\wedge Z} = S_X S_Z = S_X^* + S_Z^*$, and $E(X \wedge Z)1_{\{X\wedge Z<T\}} = \int_0^T (S_X^* + S_Z^*)$, we have:

$$\begin{aligned} P\{X < T \cap X < Z\} &= P\{X < Z\} - P\{X > T \cap X < Z\} = \\ S_X^*(0) - S_X^*(T) &= -\int_0^T dS_X^*(u) = \\ \lambda \int_0^T S_{X\wedge Z}(u)du &= \lambda \int_0^T (S_X^* + S_Z^*)(u)du. \end{aligned}$$

Since this holds for all $T > 0$,

$$\frac{dS_X^*(u)}{S_X^*(u) + S_Z^*(u)} = -\lambda.$$

The proof is concluded by substituting this into (2). The converse is an easy computation.□

It is interesting to remark that Theorem 4 is a derivation of TTT which does not appeal to the maximum likelihood estimate of λ. Indeed, put $T = \infty$ in (12), then TTT consistently estimates $1/\lambda$.

Let S_X and S_Z be strictly monotone, so that their derivatives exist everywhere. Recall that the *failure rate* of $X = -(d/dt)(\log S_X)$. We say that Z is IFR if its failure rate is increasing, and DFR if it is decreasing ([2]).

Theorem 5 *Let $X \perp Z$ with subsurvival functions S_X^* and S_Z^* respectively. Let $S_X(t) = \exp^{-\lambda t}$, and let $S_Z(t)$ be continuous and strictly monotone. If Z is DFR then*

$$\frac{S_X^*(t)}{S_X^*(0)} > \frac{S_Z^*(t)}{S_Z^*(0)};\ t > 0.$$

Similarly, if Z is IFR then:

$$\frac{S_X^*(t)}{S_X^*(0)} < \frac{S_Z^*(t)}{S_Z^*(0)};\ t > 0.$$

Proof: We prove only the first statement, as the proof of the second is similar. Let $G_Z(t) = S_Z^*(t)/S_Z^*(0)$, and $G_X(t) = S_X^*(t)/S_X^*(0)$, and put $g_Z(t) = dG_Z(t)/dt$, $g_X(t) = dG_X(t)/dt$. The assumptions on X and Z insure that these derivatives exist everywhere. Since the failure rate of X is constant, we have from (2):

$$-\lambda = \frac{dS_X^*(t)}{S_X^*(t) + S_Z^*(t)}.$$

From (3) we may write the failure rate of Z as

$$-\frac{dS_Z^*(t)}{S_X^*(t) + S_Z^*(t)} = \lambda \frac{dS_Z^*(t)}{dS_X^*(t)}. \tag{13}$$

The failure rate of Z is decreasing if and only if $g_Z(t)/g_X(t)$ is decreasing, or

$$0 > \frac{\dot{g}_Z(t)g_X(t) - \dot{g}_X(t)g_Z(t)}{g_X(t)^2}. \tag{14}$$

where $\dot{g}_Z(t) = dg_Z(t)/dt$, $\dot{g}_X(t) = dg_X(t)/dt$. Since $g_Z(0) = g_X(0) = 1$, we have $\dot{g}_Z(0) < \dot{g}_X(0)$ and for t sufficiently close to 0, we have $g_Z(t) < g_X(t)$. The set $A = \{t > 0 | g_X(t) = g_Z(t)\}$ is closed, and has a smallest element if A is non-empty. Suppose A is non-empty and that t_0 is its smallest element. Since $g_Z(t) < g_X(t)$ for $t < t_0$, we must have $\dot{g}_Z(t_0) > \dot{g}_X(y_0)$. However, since $g_Z(t_0) = g_X(t_0)$, we must have from (14) $\dot{g}_Z(t_0) < \dot{g}_X(t_0)$. Hence A is empty. $\square$.

4 Conclusions

Suppose X has an exponential distribution and that $S_Z(t)/S_Z^*(0) < S_X^*(t)/S_X^*(0)$. The data can be described either by an age-dependent censoring model, or by a random censoring model with a DFR censoring variable. The consequences of making the wrong choice in estimating the expectation of X can

be quite severe. Suppose we order the failure and censored observations as $x_1, \cdots, x_m, z_1, \cdots z_n$. If we assume $X \perp Z$ and $S_X = \exp^{-\lambda t}$, then by Theorem 4 we could consistently estimate the expectation of X as

$$\frac{\sum_{i=1}^{m} x_i + \sum_{j=1}^{n} z_j}{m}.$$

However, if we assume an age-dependent censoring model then $P\{X = x | X < Z\} = P\{X = x\}$, hence we could consistently estaimate the expectation of X as

$$\frac{\sum_{i=1}^{m} x_i}{m}.$$

In typical reliability applications $m << n$.

References

[1] O. Aalen, "Nonparametric inference in connection with multiple decrement models" *Scandinavian Journal of Statistics* 3, pp 15-27, 1976.

[2] R. E. Barlow and F. Proschan *Mathematical Theory of Reliability*, Wiley, New York, 1965.

[3] N. E. Breslow and J. Crowley, "A large sample study of the life table and product limit estimates under random censorship" *Annals of Statistics* 2,pp 437-453, 1974.

[4] D.R. Cox "The analysis of exponentially distributed life-times with two types of failure" *J. R. Statist. Soc.* B 21, pp 4141-421, 1959.

[5] H.A. David and M.L. Moeschberger *The Theory of Competing Risks* Charles Griffin and Co. London, 1978.

[6] M. Gail "A review and critique of some models used in competing risk analysis" *Biometrics* 31, pp 209-222, March 1975.

[7] A.V. Peterson Jr. "Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions", *JASA* Dec. vol. 72 Nr. 360, pp 854-858, 1977.

[8] A.V. Peterson "Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks" *Proc. Nat. Acad. Scio. USA.* vol. 73 No. 1 pp 11-13, 1976.

[9] M.E. Tarter, W.R. Freeman and L. Polissar, "Modular nonparametric subsurvival estimation" *JASA* March, vol. 85, Nr. 409, pp 29-37, 1990.

[10] A. Tsiatis, "A Nonidentifiability aspect of the problem of competing risks" *Proceedings of the National Academy of Scioences* USA, 72, No. 1 pp 20-22, 1975.